



Università di Pisa

Dipartimento di Ingegneria dell'Informazione

Corso Magistrale di Ingegneria Biomedica

Tesi di Laurea

Progettazione e implementazione di un sistema di eye
tracking stereoscopico.

Descrizione del modello e valutazione delle prestazioni in
piani bidimensionali e spazi tridimensionali.

Relatori:

Enzo Pasquale Scilingo

Prof. Antonio Lanatà

Controrelatore:

Prof. Giovanni Vozzi

Candidato:

Simone Panfilì

Anno accademico 2013/2014

Abstract

I recenti sviluppi nella tecnologia dell'eye tracking ha permesso il loro utilizzo nei più svariati ambiti. Si possono sfruttare sia in modalità diagnostica (1), in cui viene registrata la traiettoria del punto di gaze, sfruttata per effettuare studi nel campo della neurologia e della psicologia, o in modalità interattiva (1), in cui la posizione di tale punto viene impiegata per controllare e interagire con delle interfacce e degli ambienti virtuali appositamente studiati. Nonostante queste potenzialità, i sistemi di eye tracking non trovano ancora una larga diffusione per via dell'elevata preparazione richiesta per utilizzarli (2).

L'obiettivo di questo lavoro di tesi è progettare e realizzare un sistema di eye tracking stereoscopico, in grado di tracciare il punto di gaze, inteso come il punto dove si focalizza l'attenzione dell'osservatore, sia nello spazio tridimensionale sia sul piano bidimensionale dell'immagine. Per ottenere un tale risultato utilizzeremo un approccio binoculare, effettuando il tracciamento della pupilla di entrambi gli occhi. Questa tipologia di tracciamento ancora non è stata studiata a fondo e necessita di uno studio più approfondito.

Introduzione

I metodi di eye tracking sono in grado di associare matematicamente una posizione dell'occhio o della pupilla ad una posizione del punto di gaze fissato da una persona. Per punto di gaze si intende il punto nello spazio dove si incontrano gli assi visivi dei singoli occhi (2), cioè dove il soggetto sta concentrando la sua attenzione. Un sistema di eye tracking è costituito dall'insieme di componenti hardware e software in grado di tracciare per ogni istante di registrazione la posizione dell'occhio o della pupilla, rispetto alla testa, e di implementare uno dei metodi di eye tracking. Lo scopo finale di tali sistemi è quindi tracciare il posizionamento del punto di gaze soggettivo per ogni istante temporale, sul piano dell'immagine e nello spazio tridimensionale.

A noi interessa implementare uno studio del genere e contemporaneamente migliorare il tracciamento sul piano bidimensionale, già implementato nel progetto precedente.

Questo progetto si può dividere in:

- Studio delle attuali tipologie di sistemi di eye tracking
- Implementazione di un sistema di acquisizione in grado di leggere correttamente la posizione degli occhi
- Progettazione di una metodologia per il tracciamento del punto di gaze in base alla lettura effettuata dal sistema di acquisizione, metodologia che possiamo chiamare di *gaze tracking*
- Realizzazione di un setup e di una procedura

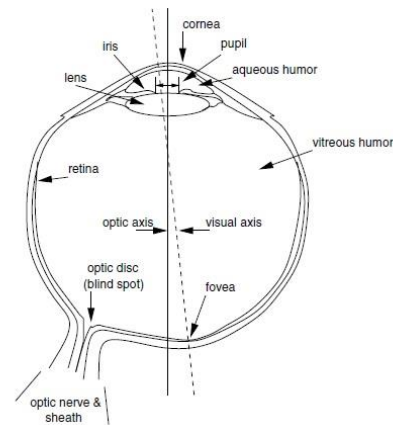
sperimentale per testare il sistema

• Analisi e discussione dei risultati

Nell'ottica di un futuro utilizzo in applicazioni reali, sarà necessario valutare la portabilità e la comodità del sistema, le prestazioni computazionali per rendere possibili applicazioni in real-time e i possibili sviluppi per migliorare tale sistema.

1. Sistemi di eye tracking

Prima di parlare dei sistemi di eye tracking, va precisato che per l'occhio si definiscono due assi:



- *asse ottico*: passante per il centro del bulbo oculare e il centro della pupilla. E' una caratteristica geometrica.
- *asse visuale*: passante per la foveola e il centro della pupilla. La foveola è la regione della retina a maggior livello di dettaglio, quindi dove si concentra l'attenzione del soggetto. E' una caratteristica funzionale ed è spostata rispetto all'asse ottico di circa 3°.

Per tracciare il punto di gaze bisogna considerare gli assi visuali dei singoli occhi.

Allo stato attuale, è possibile classificare i sistemi di eye tracking in tre grandi categorie:

- *Sistemi sensor based*: sfruttano elettrodi cutanei o lenti a contatto speciali in grado di leggere la rotazione degli occhi e di trasformarla in segnali elettrici (3).
- *Sistemi head-mounted*: sfruttano una o più videocamere in qualche modo solidali alla testa (3), tramite l'uso di un caschetto, occhiali o fasce che fissino la posizione delle videocamere relativamente agli occhi dell'osservatore.
- *Sistemi remoti*: questi usano sempre una o più videocamere e svincolano completamente il soggetto da elettrodi o sistemi indossabili.

Per quanto riguarda gli ultimi due tipi, ci sono dei processi in comune finalizzati ad analizzare l'immagine acquisita ed estrarre informazioni utili per l'eye tracking:

- *eye detection*: è la fase preliminare a tutte le successive elaborazioni. Permette di estrarre dall'immagine acquisita la regione oculare, su cui poi si andranno a effettuare i calcoli successivi.
- *pupil tracking*: una volta estratta la regione oculare, è possibile tracciare l'iride o la pupilla all'interno di questa area. Da tale rintracciamento è possibile estrarre le coordinate del centro della pupilla.
- *gaze tracking*: sono i metodi usati per tracciare il punto di gaze nello spazio o sul piano dell'immagine in base alle uscite dei blocchi precedenti. Tutti i sistemi di gaze tracking richiedono una previa calibrazione dell'hardware di acquisizione. Le tipologie di gaze tracking possono essere classificate come:

-*metodi geometrici*: cercano di ricostruire un modello geometrico dell'occhio e vanno a studiare l'andamento dell'asse ottico.

-*metodi a mappatura*: sfruttano delle mapping function, cioè dei sistemi di equazioni, lineari e non, che associano la posizione della pupilla nell'immagine ad una posizione del punto di gaze nello spazio.

Un altro aspetto importante caratterizzante i sistemi di eye tracking riguarda l'illuminazione. Questa influenza direttamente il tracking della pupilla, in quanto incide sull'accuratezza dei metodi di pupil tracking. Può essere classificata in due tipi: *illuminazione naturale* e *illuminazione a infrarossi (IR)* (3). La seconda può essere ancora suddivisa in *attiva* o *passiva*.

2. Metodologia

Il sistema scelto rientra nelle tipologie *head mounted*, a con *modello a mappatura* e ad *illuminazione infrarossa passiva*.

La strumentazione e gli algoritmi utilizzati sono stata ereditati da un progetto già realizzato presso il Centro E. Piaggio a Pisa (4) e consiste in un sistema di acquisizione indossabile, formato da una videocamera e uno specchietto usato per riflettere gli occhi nella ripresa, e di un metodo di tracciamento monoculare.

Possiamo dividere l'intero flusso del programma realizzato nei seguenti blocchi:

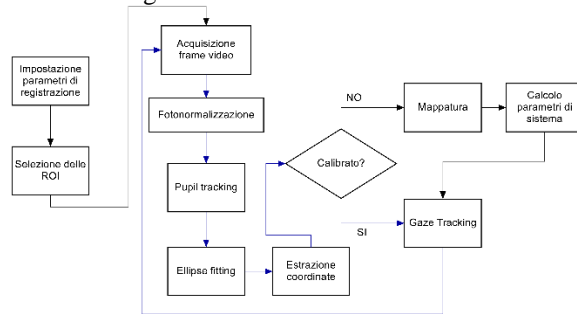


Figura 1: Flusso di programma

Dopo una fase iniziale di impostazione dei parametri per la fotonormalizzazione e la compensazione degli errori (li spiegheremo dopo), abbiamo:

-1.*Blocco di selezione delle ROI*: per ROI si intende Region of Interest, ovvero la sezione dell'immagine totale corrispondente ad uno degli occhi e su cui andremo a effettuare le successive elaborazioni. Viene effettuato manualmente all'avvio del test.

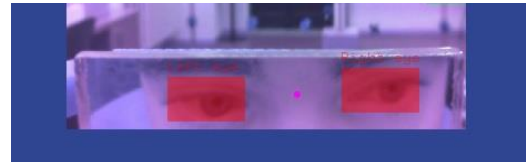


Figura 2: ROI selezionate

-2.*Blocco di fotonormalizzazione*: lo scopo delle tecniche di fotonormalizzazione consiste nel ridurre gli errori dovuti alle variazioni delle condizioni di illuminazione degli occhi, dovute a fluttuazioni dell'intensità della sorgente e ai movimenti degli occhi stessi, che ne cambiano l'esposizione. La tecnica usata in questo lavoro di tesi è chiamata SSR (Single Scale Retinex) (4).

-3.*Blocco di pupil tracking*: consiste nell'estrazione del contorno della pupilla a partire dalle informazioni

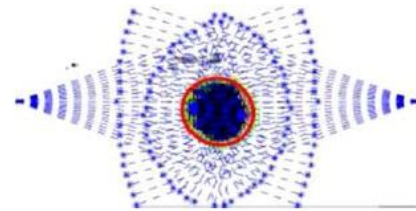


Figura 3: Rappresentazione dell'algoritmo di pupil tracking

contenute nella ROI, sfruttando il maggiore contrasto di questa sullo sfondo dovuto alla luce infrarossa ambientale.

-4.*Blocco di ellipse fitting*: dalla distribuzione di punti precedentemente estratta vengono calcolati i parametri analitici di un'ellisse. Lo stimatore è un algoritmo di tipo B2AC modificato aggiungendo un rumore di tipo gaussiano per stabilizzare la stima (5). I parametri analitici sono usati per calcolare i parametri geometrici dell'ellisse, ovvero centro, dimensione degli assi e angolazione. Il centro costituisce l'ingresso ai metodi successivi.

I blocchi finora spiegati costituiscono l'insieme di processi svolti ciclicamente dal programma, per ogni frame acquisito. Avendo scelto un metodo a mappatura bisogna dividere la sessione sperimentale in *fase di calibrazione* e *fase di registrazione*.

Prima di parlare delle due fasi sperimentali, parliamo dei sistemi di equazioni usati. Essi si riconducono alla forma generale $S = X * A$, in cui S costituisce la matrice o il vettore del punto di gaze in uscita, X la matrice o vettore

di ingresso dato dai centri delle ellissi fittate ed elaborate secondo diverse formulazioni, A la matrice di sistema descrivente la funzione di mappatura dello spazio, contenente i coefficienti delle formule usate.

Lo scopo della fase di calibrazione è costruire delle matrici S e X, in cui per ogni ingresso X_i è associata una uscita S_i , e quindi calcolare la matrice di sistema A con la formula $A = (X^T X)^{-1} X^T S$, ovvero la pseudo-inversa di Moore-Penrose (6).

Passiamo ora a parlare degli algoritmi usati:

- **Sistema Monoculare (SM)**: sfrutta le informazioni di un solo occhio e può essere usato solo per il tracking bidimensionale. E' il metodo usato anche nel progetto di partenza ed è stato usato come parametro di controllo per valutare gli altri algoritmi dal punto di vista bidimensionale. Associa gli ingressi x_{ei}, y_{ei} relativi al centro pupilla dell'occhio sinistro all'uscita x_{si}, y_{si} del punto di gaze sullo schermo:

$$\begin{aligned} x_{si} &= a_1 + a_2 x_{ei} + a_3 y_{ei} + a_4 x_{ei} y_{ei} + a_5 x_{ei}^2 + a_6 y_{ei}^2 \\ y_{si} &= b_1 + b_2 x_{ei} + b_3 y_{ei} + b_4 x_{ei} y_{ei} + b_5 x_{ei}^2 + b_6 y_{ei}^2 \end{aligned}$$

- **Sistema Binoculare con Mappatura Lineare (SBML)**: associa gli ingressi x_{li}, y_{li} e x_{ri}, y_{ri} dei due occhi all'uscita x_{si}, y_{si} secondo il sistema di equazioni:

$$\begin{aligned} x_{si} &= a_1 + a_2 x_{li} + a_3 y_{li} + a_4 x_{ri} + a_5 y_{ri} \\ y_{si} &= b_1 + b_2 x_{li} + b_3 y_{li} + b_4 x_{ri} + b_5 y_{ri} \end{aligned}$$

dove x_{si}, y_{si} sono le coordinate cartesiane reali dei punti di calibrazione. Nel bidimensionale consideriamo le prime due equazioni e x_{si}, y_{si} sono su schermo.

- **Sistema Binoculare con Mappatura Quadratica (SBMQ)**: associa gli ingressi x_{li}, y_{li} e x_{ri}, y_{ri} dei due occhi all'uscita x_{si}, y_{si} secondo un sistema di equazioni quadratiche analoghe al sistema a mappatura lineare.

- **Sistema Binoculare con Sistemi Monoculari Indipendenti (SBSMI)**: è un metodo ibrido che sfrutta due metodi monoculari, uno per occhio, calcola i punti di gaze bidimensionali destro e sinistro, ne prende la media e la corregge con uno spostamento pari a $B = \frac{-D_I * d}{D_S}$, dove D_I è la distanza su schermo dei punti di gaze, d è la distanza del baricentro dei centri pupilla dal baricentro degli occhi e D_S è la distanza istantanea dei centri pupilla. Questo ci consente di calcolare il punto di gaze bidimensionale del metodo binoculare.

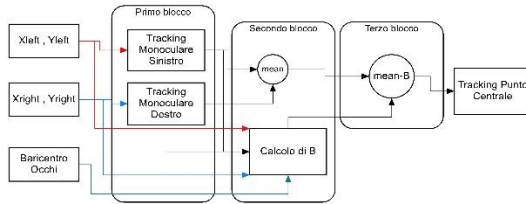


Figura 4: Schema del metodo binoculare

La profondità del punto di gaze, cioè la distanza dall'osservatore, viene calcolata come

$$D_{si} = c_1 + c_2 d_i + c_3 d_i^2$$

in maniera simile ai sistemi già spiegati, dove d rappresenta la distanza inter-pupillare.

3. Strumentazione

Il sistema di acquisizione, chiamato HATCAM, è costituito da una videocamera, uno specchietto e il telaio su cui sono montati, più le fasce elastiche per indossarlo. La videocamera montata è una Microsoft LifeCam HD-3000 (7), con una modifica fondamentale: le è stato tolto il filtro infrarosso per fare in modo di acquisire la banda infrarossa della luce naturale. La risoluzione impiegata si attesta sui 720x576 pixel.

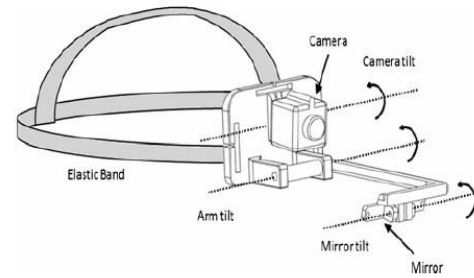


Figura 5: Sistema HATCAM

Il setup sperimentale è una griglia composta da 25 punti disposti su un primo perpendicolare all'asse visivo della camera distante 350 cm dal soggetto. Ci sono altri 6 bersagli disposti posteriormente al primo piano e leggermente scostati, in modo da rimanere osservabili durante l'esperimento (8), lontani 425 cm e 500 cm.

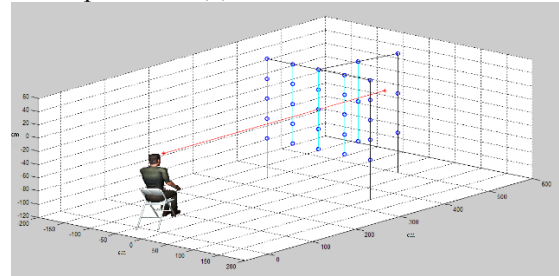


Figura 6: Schema del setup sperimentale

4. Esperimento

Sono state effettuate due sedute sperimentali, registrando la prova di un totale di 13 soggetti, senza problemi alla vista per non influenzare i risultati.

La procedura sperimentale consiste nella fase di calibrazione, in cui il soggetto fissa in successione i 25 punti per la calibrazione bidimensionale e in seguito i 3 punti per la calibrazione della profondità.

Finita la calibrazione, si è proceduto con la registrazione della prova, in cui si chiede al partecipante di fissare in successione 9 dei bersagli per valutare gli algoritmi

bidimensionali e poi altri 9 per valutare quelli tridimensionali.

5. Risultati

Data la presenza di outliers dovuti ai blink e agli spostamenti della camera rispetto al soggetto, abbiamo usato come parametro la mediana, in modo da eliminare il loro contributo. Inoltre dal test di Shapiro-Wilk risulta che i dati non sono distribuiti normalmente, perciò usiamo dei test non parametrici.

Valutazione nel caso bidimensionale:

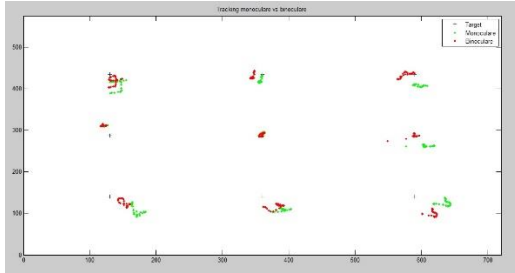


Figura 7: Grafico dei punti di fissazione

Siamo andati a valutare l'errore angolare sia orizzontale (ϵ_x) sia verticale (ϵ_y) nel tracciamento del punto di gaze sul piano dell'immagine.

		SM	SBML	SBMQ	SBSMI
ϵ_x	75° percentile	2.2177°	1.5195°	2.1257°	1.8835°
	mediana	1.1108°	0.8322°	0.9603°	0.9066°
ϵ_y	75° percentile	2.9699°	2.6525°	2.5080°	3.0470°
	mediana	1.6777°	1.3727°	1.2682°	1.3536°

Dall'analisi statistica effettuata, i nuovi sistemi risultano leggermente migliori rispetto al metodo monocular, specialmente il SBML, ma statisticamente non si sono trovate differenze significative. Riportando i risultati dei test di Wilcoxon appaiati, seguendo l'ipotesi che il metodo monocular restituisca un errore maggiore rispetto agli altri:

H0 del test di Wilcoxon		SM > SBML	SM > SBMQ	SM > SBSMI
ϵ_x	p-value	0.97988	0.74751	0.73085
ϵ_y	p-value	0.8908	0.92448	0.67779

Valutazione nel caso tridimensionale:

Dall'analisi sulla prestazione delle varianti tridimensionali degli algoritmi sul piano bidimensionale sono emersi risultati analoghi ai precedenti.

Per valutare le prestazioni nel tridimensionale siamo andati a misurare la differenza fra profondità stimata e profondità del punto reale.

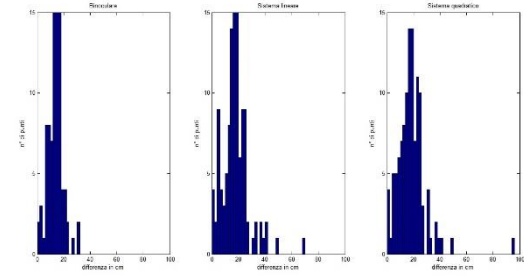


Figura 8: Istogrammi relativi a ϵ_D

		SBML	SBMQ	SBSMI
ϵ_D e percentuale sulla profondità reale	75° percentile	23.23cm	23.40cm	16.53cm
	mediana	17.21cm	17.86cm	14.76cm
		4.05 %	4.20 %	3.47 %

Effettuando il test di Kruskal-Wallis con correzione di Bonferroni:

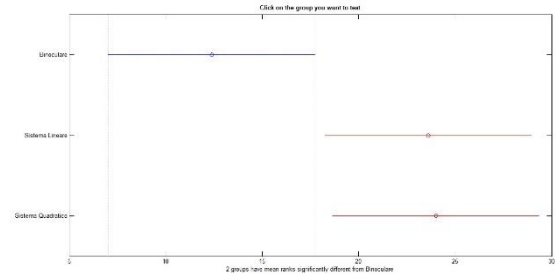


Figura 9: Test di Kruskal-Wallis per la differenza nella stima della profondità

Tempi di computazione: sono stati misurati anche i tempi di elaborazione per ogni frame acquisito e risultano 84.6 ± 9.1 ms. Il frame rate del sistema risulta quindi di circa 11.82 ± 1.35 fps.

6. Conclusioni e sviluppi futuri

Analizzando i risultati per entrambi i casi di studio, possiamo dire che i sistemi SBML e SBMQ sono più accurati nel tracciamento del punto di gaze sul piano bidimensionale rispetto al SBSMI. Tali sistemi possono ancora essere studiati e approfonditi per cercare di migliorare la loro accuratezza.

Per quanto riguarda il tracciamento tridimensionale, la stima della profondità effettuata dal SBSMI risulta migliore degli altri sistemi e dimostra un'accuratezza soddisfacente.

Ci sono molti possibili sviluppi nel prossimo futuro. Sicuramente, la prospettiva più interessante è integrare i due metodi migliori, sistema lineare e calcolo della profondità, per ottenere un tracciamento accurato sia in 2D che in 3D. Dal punto di vista software, si potrebbe incrementare il frame rate del sistema utilizzando tecniche di multi-threading. Infine la prospettiva più interessante

riguarda uno studio più approfondito del comportamento della profondità del punto di gaze in base alla nostra lettura della posizione degli occhi, dando la possibilità di un tracciamento più accurato in futuro. Un sistema del genere può essere usato in applicazioni in real-time, fornendo informazioni aggiuntive sul dove e cosa sta guardando l'osservatore, ampliando le possibilità di utilizzo dei sistemi di eye tracking.

References

1. **Amer Al-Rahayfeh & Miad Faecipur.** *Eye Tracking and Head Movement Detection: A State-of-Art Survey.*
2. **Andrew T. Duchowski.** *Eye Tracking Methodology Theory And Practice 2nd Edition.*
3. **Dan Witzner Hansen, Qiang Ji.** *In the Eye of the Beholder: A Survey of Models for Eyes and Gaze.*
4. **Antonino Armato, Antonio Lanata`, Enzo Pasquale Scilingo.** *Comparitive study on photometric normalization algorithms for an innovative, robust and real-time eye gaze tracker.*
5. **Eliseo Stefano Maini.** *Robust Ellipse-Specific Fitting for Real-Time Machine Vision.*
6. **Luigi Landini, Fabio Gibiino.** *Metodo dei minimi quadrati con regolarizzazione e applicazioni.*
7. **Microsoft Corporation.** *Datasheet Microsoft LifeCam HD-3000.*
8. **Andrew T. Duchowski, Donald H. Housey, Jordan Gestringy, Robert Congdonz.** *Comparing Estimated Gaze Depth in Virtual and Physical Environments.*